

**Spezialschulteil für Mathematik /
Naturwissenschaften / Informatik**

am Albert-Schweitzer-Gymnasium in Erfurt

KLASSENSTUFE: 12

SCHULJAHR: 2022-2024

Pneumonia Detection mithilfe von Convolutional Neural Networks

Betreuer:

Herr Süpke

Herr Golchert

Frau Dr. Moor

Unterstützung:

Stiftung Jugend forscht e. V.

Deutsche Unfallkasse

Universitätsklinikum Jena

Verfasser:

Dustin Marggraff

Nico Lentsch

Marvin Heyne

20.06.2024 / Erfurt

Inhaltsverzeichnis

1	Einleitung	4
2	Hintergrund und theoretische Grundlagen	5
2.1	Convolutional Neural Networks	5
2.1.1	Verbreitete CNN-Architekturen	6
2.1.2	Bild-Segmentierung	8
3	Materialien	10
3.1	Verwendete Module	10
3.2	Datensätze	11
4	Vorgehensweise	11
4.1	Erstellen eines einfachen CNN	12
4.2	Untersuchungen zur Klassifizierung	12
4.3	Untersuchung zur Object Detection	13
4.4	Untersuchung 2-teiliges Netz	13
4.5	Untersuchung kombinierter Netzwerke	13
4.6	Integration der KI in den medizinischen Workflow	16
5	Literaturverzeichnis	17

1 Einleitung

Pneumonien, auch als Lungenentzündungen bekannt, sind aufgrund ihrer hohen Prävalenz und globalen Verbreitung seit Jahrhunderten eine bedeutende gesundheitliche Herausforderung. Trotz umfangreicher Forschung zählt die Pneumonie weiterhin zu den häufigsten Lungenerkrankungen weltweit. Allein in Deutschland sterben jährlich zwischen 30.000 und 40.000 Menschen an den Folgen dieser Erkrankung.

Ein wesentlicher Faktor für die hohe Mortalitätsrate ist die Komplexität der Diagnose, bei der im Durchschnitt etwa 25 % der betroffenen Patienten nicht korrekt erkannt werden. Dies trägt nicht nur zur genannten Sterblichkeitsrate bei, sondern führt auch zu Langzeitschäden und weiteren gesundheitlichen Komplikationen.

Während typische Symptome wie Husten, Atemnot und Fieber zwar häufig im Rahmen einer Anamnese festgestellt werden, ermöglichen sie meist keine präzise Diagnose. Aus diesem Grund greifen Ärzte, insbesondere bei schwerwiegenden Verläufen, auf Röntgenbilder zurück. Die Interpretation dieser Bilder stellt jedoch eine Herausforderung dar und führt nicht immer zu einer eindeutigen Diagnose.

Diese Arbeit widmet sich der Entwicklung eines KI-Systems, das auf Convolutional Neural Networks (CNN) basiert und zur automatisierten Erkennung von Pneumonien auf Röntgenbildern eingesetzt wird. Ziel der Forschung ist es, die Diagnosegenauigkeit zu erhöhen und Fehldiagnosen zu minimieren, um medizinisches Fachpersonal bei der Beurteilung von Patienten zu unterstützen. Dabei ist das in dieser Arbeit entwickelte Tool als ergänzendes und beratendes Hilfsmittel gedacht, das Ärzten im diagnostischen Prozess zur Verfügung steht.

Angesichts der hohen Prävalenz und der signifikanten Todesrate wird in dieser Arbeit die Pneumonie als exemplarische Lungenerkrankung untersucht. Verschiedene Ansätze und KI-Modelle der Computer Vision werden für diese Aufgabe optimiert und miteinander verglichen. Zudem wird der Einfluss von patientenbezogenen Metadaten auf die Diagnosegenauigkeit untersucht. Ein weiterer Forschungsaspekt ist die Integration dieser Anwendung in den medizinischen Workflow.

Für das Training der entwickelten KIs wurden Datensätze verwendet, die von internationalen medizinischen Einrichtungen bereitgestellt und von Experten annotiert wurden. Die Verifizierung und Integration des Systems in die medizinische Praxis erfolgt in enger Zusammenarbeit mit dem Universitätsklinikum Jena.

Abschließend möchten wir unseren Dank an die Betreuer dieser Arbeit, Johannes F. Süpke, Dr. Marion Moor und Herrn Golchert, sowie an alle beteiligten Partner aussprechen. Auch den Finanziellen Unterstützung durch die Deutsche Unfallkasse, sowie den Jugend Forscht Verband möchten wir danken.

2 Hintergrund und theoretische Grundlagen

Lungenentzündungen (Pneumonie) sind aufgrund ihrer Prävalenz und ihrer geografisch weitreichenden Verbreitung eine seit Jahrhunderten bekannte und erforschte Krankheit. Bis heute ist es jedoch nicht möglich, sie aufgrund ihrer zahlreichen Ursachen, vollständig zu präventieren und teilweise auch nicht sie zu heilen. Zu den bekannten Ursachen zählen Erreger, wie Viren, Bakterien und Pilze oder auch chemische Reize, sowie Feinstaub. Dies führt häufig zu Entzündungen der Bronchien, des Lungenzwischengewebes oder der Alveolen. Diese können sich mit Eiter oder Schleim füllen, was zu Symptomen wie Husten, Atemnot, Fieber oder Auswurf, bei schweren bzw. tödlichen Verläufen auch zu einer Sepsis führt.

Die Diagnose erfolgt meist durch eine Anamnese, bei der die gesundheitliche Vorgeschichte und anderen Merkmalen wie Rauchen oder dem BMI untersucht werden. Diese ist jedoch nur begrenzt aussagekräftig. Deshalb greifen Ärzte, vor allem bei schwereren Verläufen, auf das Röntgen-Thorax, also ein Röntgenbild des Brustkorbes, zurück. Die Verwendung von MRTs und CTs zur Diagnose ist auf Grund der geringeren Genauigkeit und kompliziertere Durchführung, seltener.

Auf Röntgenbildern achten Ärzte auf veränderte Strukturen. Leichte erhellte Bereiche sind bei Konsolidierung, also Ansammlung von Flüssigkeit im Gewebe oder verdichtetem Gewebe, erkennbar, während es zu getrübten Bereichen bei beispielsweise Infiltraten kommt. Eine frühe Diagnose ist entscheidend, da sie dem Körper mehr Zeit gibt den Erreger zu bekämpfen. Man kann ihn dabei, mithilfe von Virostatika, Antibiotika oder symptomlindernden Medikamenten unterstützen. Eine verspätete Behandlung kann zu schwereren Verläufen, sowie zu dauerhafte Schädigungen führen und sogar andere Organe in Mitleidenschaft ziehen, was die Behandlung erheblich erschwert.

2.1 Convolutional Neural Networks

In der Erkennung von Lungenentzündungen auf Röntgenbildern bieten Computer-Aided Diagnosis (CAD)-Systeme wertvolle Unterstützung. Ein vielversprechender Ansatz dafür sind Convolutional Neural Networks (CNN). Sie bieten durch ihren speziellen Aufbau besondere Qualifikationen für die Verarbeitung von strukturierten Datenarrays, wie Bild- und Audiodateien, was sie zu den am weitesten verbreiteten Netzstrukturen im Bereich der Computer Vision macht. [28] [31] Neuronale Netze, die dem menschlichen Gehirn nachempfunden sind, bestehen aus Knoten, sogenannten Neuronen, die eingehende Informationen summieren und diese durch eine nichtlineare Aktivierungsfunktion weiterverarbeiten. Diese Funktion ahmt die Schwellwerte im Gehirn nach. Durch Verbindung mehrerer Neuronen entsteht ein Neuronales Netz, welches in der Lage ist logische Probleme zu bearbeiten. Dabei sind die Verbindungen gewichtet, wodurch die Informationen verändert und priorisiert werden können. Meist sind die Neuronen eines Netzes auf Schichten (Layers) angeordnet, welche sich in Aufbau und Funktion unterscheiden.

Convolutional Neural Networks (CNNs) sind eine Spezialisierung der Neuronalen Netze und von der Schinde des menschlichen Gehirns inspiriert. Sie ermöglichen die Extraktion und Klassifizierung von Merkmalen und Strukturen aus Bildern und zeichnen sich durch hohe Speicher- und Laufzeiteffizienz aus. Ein CNN ist meist aus drei unterschiedlichen Schichtarten aufgebaut, welche speziell für die Bildanalyse entwickelt wurden (siehe Abb. 1).

Zu Beginn eines CNNs steht meist eine Kombination aus Pooling und Convolutional Layer (Conv-Layer). Die Convolutional Layers, auch Faltungsschichten genannt, verwenden Filter¹, die in der Lage

¹Filter sind Matrizen mit Zahlen, welche mit einer bestimmten Schrittweite über die Eingabe bewegt, wobei die überdeckten Bereiche jeweils mit dem Filter multipliziert werden. Die Summe der Produkte bildet ein Pixel der Ausgabe

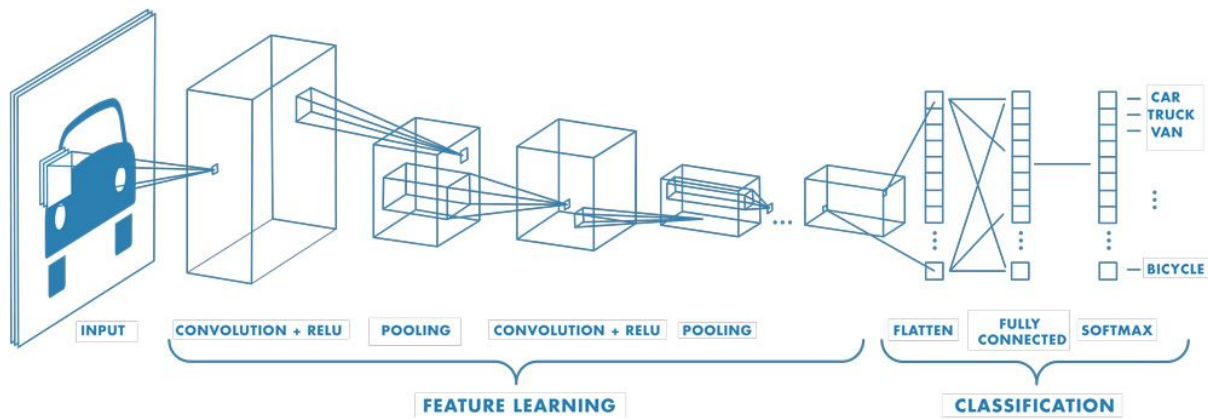


Abbildung 1: Aufbau eines CNN [25]

sind, Merkmale wie Ecken oder Kanten hervorzuheben. Mehrere aufeinanderfolgende Faltungsschichten können zunehmend komplexere Muster erkennen.

Durch diese Struktur können Merkmale in einer zweidimensionalen Eingabe unabhängig von ihrer Position im Bild erkannt werden. [3] [19] Darauf folgende Pooling Layer reduzieren die Datenmenge, indem sie einen Bildbereich (fester Größe) zu einem Wert zusammenfassen. Ein häufig eingesetztes Verfahren ist das Max-Pooling, bei dem der höchste Wert eines Bereichs in den Ausgabepixel übernommen wird. Dadurch wird die Rechen- und Speicherleistung des Netzes verringert und irrelevante Informationen sowie Rauschen entfernt. [30]

Die letzten Schichten werden stets von mindestens einem Fully-Connected-Layer (Dense-Layer) gebildet. Die Eingabe dieser Schicht ist eindimensional weshalb die vorige zweidimensionalen Ausgaben von z.B: Conv-Layer durch aneinanderreihen der einzelnen Bildzeilen geflattet wird (Flatten-Layer). Jedes Neuron einer solchen Schicht ist jeweils mit allen Neuronen der vorigen Schicht über eine gewichtete Verbindung verknüpft. Die durch die Gewichte veränderte Eingaben bzw. In dieser Schicht werden die vorher herausgefilterten Merkmale klassifiziert. [33]

Beim Training eines CNNs werden dessen Parameter wie Gewichte oder Filter initialisiert und anschließend durch ein Lernverfahren an die spezifische Aufgabe angepasst. In dieser Arbeit wird die Methode des überwachten Lernen (Supervised-learning) angewendet. Bei dieser wird ein Datensatz benötigt, welcher Trainingsdaten und dazugehörige Label (Lösungen) enthält. Während der Trainingsphase wird die Ausgabe des Netzes mit den Labels verglichen, und die Parameter werden in der Backpropagation mithilfe des Gradient Descent-Verfahrens optimiert. Die Anpassung der Parameter hängt dabei von der Größe des Fehlers, dem verwendeten Optimizer und der Lernrate ab. [18]

2.1.1 Verbreitete CNN-Architekturen

Convolutional Neural Networks wurden in den 90ziger Jahren von Yann LeCun vorgeschlagen. Bis heute folgen nahezu alle CNNs dem damals vorgeschlagenen Design. Während anfangs noch eine simple Steigerung der Anzahl der Conv-Layer zu besseren Erfolgen führten, entstanden neue innovative Ansätze um die Genauigkeit aber auch Effizienz zu steigern. Diese Netze finden häufig Anwendung bei der Bildklassifizierung. Diese versuchte ein Eingabebild einer bestimmten Klasse z.B: „Pneumonie“ oder „Gesund“ zuzuordnen.

Die von Lecun vorgeschlagene Standardarchitektur für CNNs ähnelt der in Abbildung 1 und wurde unter der Bezeichnung LeNet für die Erkennung Handgeschriebener Zahlen bekannt. [8]

Mit steigender Rechenleistung, stieg auch die Größe der Netzwerke. 2012 gelang Alex Krizhevsky

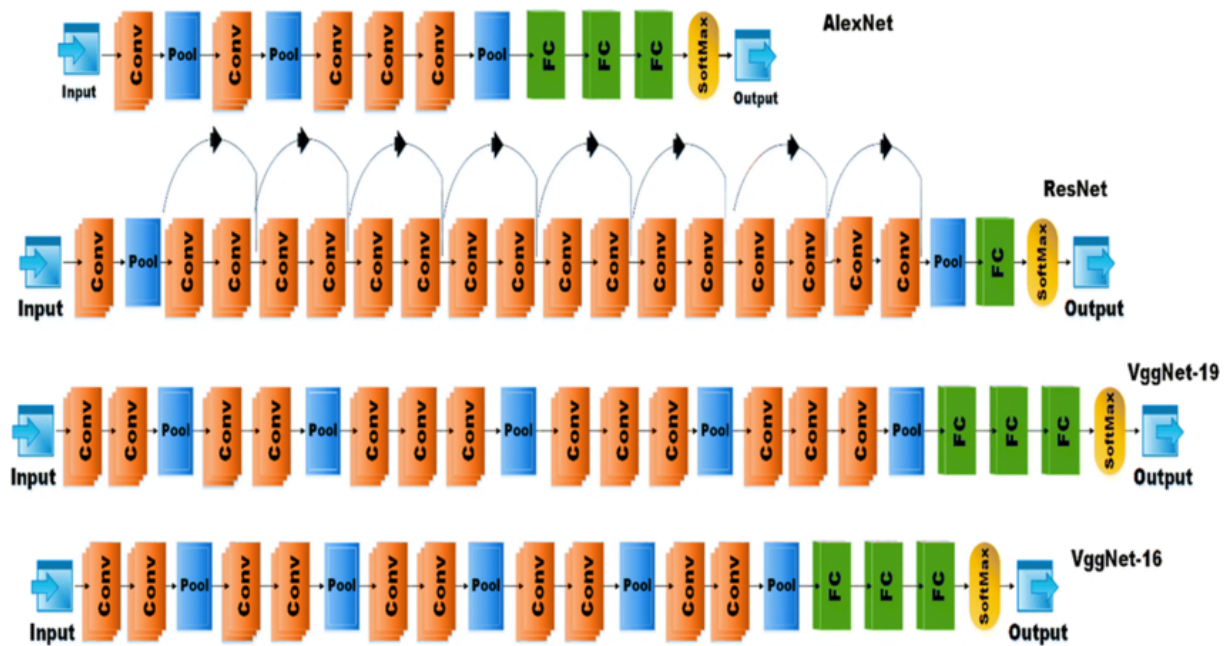


Abbildung 2: Vergleich von AlexNet, ResNet, VGG-19 und VGG-16 [32]

et al. ein markanter Durchbruch mit seiner Architektur AlexNet bei der ImageNet ²Competition. Die Architektur ist ähnlich zum Aufbau von LeNet besitzt jedoch statt 60 tausend 60 Millionen Parameter und ist somit deutlich größer. [7]

Mit insgesamt 138 Millionen Parameter wurde 2 Jahre später VGG16 vorgestellt, welches abermals Größer als AlexNet ist (siehe Abbildung 2). Das Besondere ist, dass alle Convolutional Kernels die Größe von 3x3 und alle MaxPooling-Kernels 2x2 mit Stride 2 besitzen. Mit dieser Kernel Größe (3x3) lassen sich alle andern Kernel Größen nachbauen. Für z.B.: einen 5x5 Kernel benötigt man 2 3x3 Convolutionen hintereinander, jedoch statt $5 \times 5 = 25$ nur $2 \times 3 \times 3 = 18$ Parameter.[13] Ein neuer Ansatz wurde 2014 im Inception (GoogleLeNet) vorgestellt. Diese Struktur besteht aus mehreren Blöcken, welche mehrerer paralleler Pfade enthalten. Jeder Pfad wiederum enthält eine Aneinanderreihung mehrerer Conv-Layer. Am Ende werden die Ausgaben der Pfade zu einer Ausgabe zusammengefasst. Somit ist es möglich Objekte verschiedener Größen zu betrachten (siehe 3), in dem auf verschiedenen Pfaden 1x1, 3x3, und 5x5 Convolution ausgeführt werden kann. [14]

Eine bedeutende Innovation für Convolutional Neural Networks (CNNs) wurde 2015 von Kaiming He, Xiangyu Zhang, Shaoqing Ren und Jian Sun mit der Einführung der ResNet-Architektur (Residual Network) vorgestellt. Bis dahin hatten neuronale Netze oft das Problem, dass sie keine einfache Lösung finden konnten. Zudem führte das Hinzufügen von Schichten nicht zu einer erwarteten Steigerung der Genauigkeit, sondern zu deren Abnahme. Ein wesentlicher Grund dafür ist das Vanishing Gradient Problem, bei dem Parameter in tieferen Schichten nur sehr langsam angepasst werden.

Als Lösung schlugen die Autoren sogenannte Residual-Verknüpfungen vor (siehe Abb. ??), die als Shortcut-Verbindungen fungieren. Diese Verbindungen ermöglichen es tiefen neuronalen Netzwerken, effizienter und stabiler zu trainieren, indem sie das Problem der verschwindenden Gradienten umgehen. Ein ResNet besteht aus Residual-Blöcken, die eine Eingabe x verarbeiten, diese durch mehrere Conv-Layer leiten und dann die Ausgabe $F(x) + x$ zurückgeben. Dabei werden zwei Arten von Verbindungen unterschieden: Erstens die Identitäts-Shortcut-Verbindung, bei der die Eingabe x unverändert zur Ausgabe addiert wird, und zweitens die Projektions-Shortcut-Verbindung, bei der sich die Dimensionen von

²ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) war ein bedeutender Wettbewerb im Bereich der Computer Vision. Die Netze wurden dabei auf einem Datensatz von über 10 Millionen hangeschriebenen Bildern trainiert.

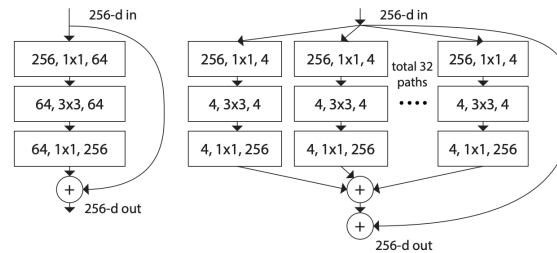


Figure 1. **Left:** A block of ResNet [14]. **Right:** A block of ResNeXt with cardinality = 32, with roughly the same complexity. A layer is shown as (# in channels, filter size, # out channels).

Abbildung 3: Beispiel eines ResNext-Blockes

Eingabe x und Ausgabe $F(x)$ unterscheiden, weshalb die Eingabe x transformiert werden muss.

Diese Verbindungen gewährleisten nicht nur, dass die Gradienten in tieferen Schichten vergleichsweise stark bleiben, sondern auch, dass das Netzwerk, da jeder Residual-Block die gleiche Abbildung lernen kann, in der Lage ist, eine einfache und effektive Lösung für ein Problem zu finden. Es gibt verschiedene ResNet-konfigurationen, welche sich in ihrer Anzahl an Convolutional Schichten unterscheiden. Diese Struktur ist bis heute, auch als NTeil anderer Netzarchitekturen, im Einsatz. In dieser Arbeit werden z.B. ResNet50 und Resnet101 als Backbone-Strukturen verwendet.[5]

Eine Verbindung von Interception und ResNet findet sich in der ResNeXt Architektur. Sie fügt den Interception-Blöcken zusätzlich eine Residuale Verbindung hinzu. Die Inputchannel eines Blockes werden dabei auf die verschiedenen Pfade verteilt. Dadurch ist es möglich die Convolution der Pfade gleichzeitig auf mehreren GPUs auszuführen. Des Weiteren führen die Gruppen (Convolution auf den verschiedenen Pfaden) zu einer Spezialisierung, nach der z.B: ein Gruppe die Schwarz und Weiß formen untersucht und eine andere die Farben des Bildes. Erstmals wurde diese Gruppen Convolution im Alex-Net Paper vorgeschlagen.[16]

DenseNet führt die Idee von Rückführenden Verbindungen noch weiter. Hier kann jede feature Map alle vorhergehenden feature Maps (Ausgabe von Conv-Layer) einsehen. Um dies zu erreichen wird jede Ausgabe eines vergangenen Conv-Layer auf die Ausgabe des aktuellen Conv-Layer draufaddiert. Die Parametergröße wird durch die geringe Ausgabegröße (Channel-dimensions) von z.B. 12 Filter, verringert. ein DenseNet besteht meist aus mehreren dieser Dense Blocks zwischen denen sich jeweils noch ein Conv und Pooling-Layer befinden. Durch diesen Aufbau erreicht die Struktur bessere Resultate als ResNet bei einer geringeren Komplexität.[24][6]

Dies sind alle gängigen Netzstrukturen, welche bis heute noch Anwendung in vielen Bereichen der Computer-Vision finden. Diese Netzwerke wurden zur Image-Klassifizierung geschaffen, also um eine Klasse auf einem Bild festzustellen. Jedoch finden sie auch Anwendung als Teilstrukturen in Netzarten der Image-Segmentierung.

2.1.2 Bild-Segmentierung

Neben der Klassifizierung spielt die Segmentierung eine entscheidende Rolle in der Computer Vision. Dabei wird zwischen semantischer Segmentierung und Instanz-Segmentierung unterschieden. Bei der semantischen Segmentierung wird versucht, alle Objekte einer bestimmten Klasse im Bild zu identifizieren und vom Hintergrund zu trennen. Die Instanz-Segmentierung geht einen Schritt weiter, indem sie zusätzlich verschiedene Objekte derselben Klasse voneinander differenziert. Dadurch wird nicht nur festgelegt, ob ein Pixel zu einer Person oder zum Hintergrund gehört, sondern auch zu welcher spezifi-

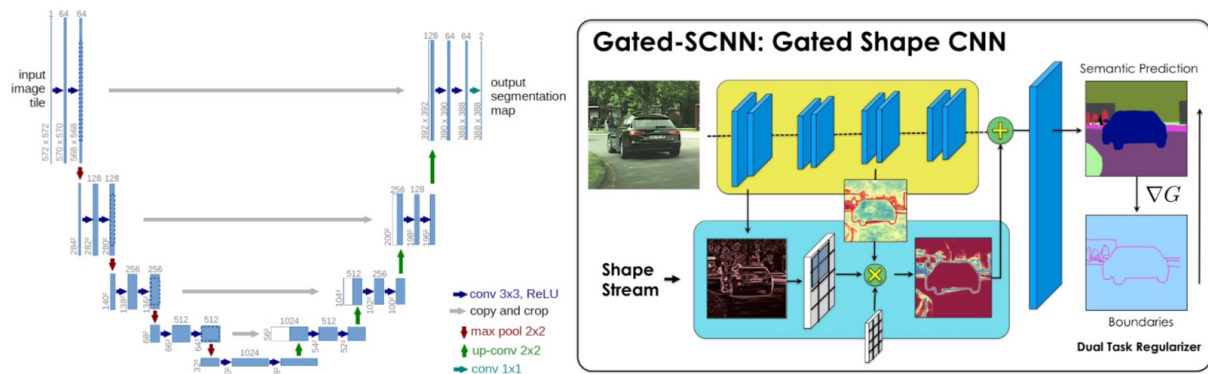


Abbildung 4: Visualisierung der Architektur von UNet[12](links) und GSCNN[15] (rechts)

schen Person (z.B. Person 1, Person 2, usw.). Diese Verfahren ermöglichen auch die Erkennung mehrerer Klassen in einem Bild.

Ein bekanntes Netzwerk zur semantischen Segmentierung ist das U-Net. Diese Art von Convolutional Neural Network (CNN) ist durch eine spezielle Architektur gekennzeichnet. Es beginnt mit sogenannten Conv-Blöcken, die die Bildgröße nicht verändern, gefolgt von einer Pooling-Schicht, die die Bildgröße halbiert. Der Aufbau des Downsampling-Pfades ähnelt dem eines VGG-Netzwerks. In diesem Pfad wird das Bild schrittweise verkleinert, wobei immer komplexere Merkmale extrahiert werden.

Nach der Verkleinerung des Bildes folgt der Upsampling-Pfad. Dieser beginnt mit einer Upsampling-Schicht, die die Bildgröße verdoppelt, gefolgt von einem Conv-Block, der die Bildgröße beibehält. Eine Besonderheit des U-Net besteht darin, dass äquivalente Conv-Blöcke aus dem Downsampling- und Upsampling-Pfad miteinander verbunden sind. Die Ausgabe des Downsampling-Pfades wird dabei zu den entsprechenden Conv-Blöcken des Upsampling-Pfades hinzugefügt. Dadurch können Merkmale verschiedener Größen und Komplexitäten erkannt werden (siehe Abb. 4). Am Ende des Netzwerks steht oft eine Sigmoid-Funktion, die die Merkmalskarte in eine binäre Maske umwandelt, wobei die Pixel der erkannten Objekte aktiviert und die des Hintergrunds inaktiv bleiben. Es gibt eine Vielzahl von Erweiterungen dieser Architektur wie UNet++[17], VUNet [?] oder ResUNet[2]. Sie fügen unter anderem der Standardstruktur Residuale oder Attention-Verbindungen hinzu.

Gated Shape CNNs sind eine weitere Form von Convolutional Neural Networks, die darauf ausgelegt sind, die geometrischen Formen von Objekten in Bildern effektiver zu erkennen und zu verarbeiten. Diese Netzwerke nutzen "Gates", also spezielle Aktivierungsmechanismen, die es dem Netzwerk ermöglichen, Informationen über die Form von Objekten selektiv zu filtern und hervorzuheben. Durch die gezielte Verarbeitung von Forminformationen können Gated Shape CNNs besonders gut bei der Erkennung von Objekten mit klar definierten Konturen eingesetzt werden, was sie in Anwendungen wie der medizinischen Bildverarbeitung oder der automatisierten Fahrzeugerkennung besonders nützlich macht. Sie nutzen hierfür eine beliebige Tiefe Standardstruktur wie ResNet oder DensNet als Backbone Network auf dem ersten Pfad. Parallel zu diesem fügen die Entwickler einen separaten Formen Pfad (Shape Stream) hinzu, der die Grenzen zwischen Objekt und Hintergrund erlernt. Er wird dabei über Gates mit Wissen aus dem 1. Pfad (Classical Stream). Dadurch werden schärfere Vorhersagen zu der Separation von Objekt und Hintergrund möglich.[15]

DeepLab ist ein weiteres fortschrittliches Modell, das auf der semantischen Segmentierung spezialisiert ist. Es verwendet Atrous Convolution (auch dilated convolution genannt), um den Empfangsbereich der Convolutional Layers zu erweitern, ohne die räumliche Auflösung der Merkmalskarten zu verrin-

gern. Dies ermöglicht es DeepLab, kontextuelle Informationen aus größeren Bereichen des Bildes zu erfassen und feine Details präzise zu segmentieren. DeepLab integriert außerdem eine Fully Connected Conditional Random Field (CRF) zur Nachbearbeitung, um die Segmentierungsergebnisse weiter zu verfeinern und klare Grenzen zwischen Objekten zu erzielen. Diese Kombination macht DeepLab besonders leistungsfähig bei der Segmentierung in komplexen Szenen.[1]

Für die Instance-Segmentation wurden fortschrittlichere Modelle entwickelt. Man unterscheidet in Single- und Two-Shot detectors. Single shot Detectoren werden für die schnelle Verarbeitung eingesetzt, da sie das Bild nur einmal verarbeiten. Dazu werden bei Initialisierung das Bild in bestimmte Bereiche. Dabei wird bestimmt ob diese Objekte enthalten und wie die Grenzen angepasst werden müssen um die Objekte ganz zu umfassen. Beispiel hierfür sind YOLO[11], SSD[10], RetinaNet[9].

Für eine höhere Genauigkeit werden zweistufige (two-shot) Detectoren verwendet. Beispielhaft hierfür ist die RCNN Familie. In dieser Arbeit werden dabei Mask R-CNNs verwendet. Mask R-CNN erweitert das Konzept des Region-based Convolutional Neural Network (R-CNN) und seines Nachfolgers Faster R-CNN, um neben der Erkennung und Klassifizierung von Objekten auch deren pixelgenaue Segmentierung durchzuführen. Mask R-CNN fügt dem bestehenden Framework eine zusätzliche "Mask Branch" hinzu, die für jede Region of Interest (RoI) eine binäre Maske erzeugt. Diese Maske gibt an, welche Pixel innerhalb der RoI zu dem erkannten Objekt gehören. Mask R-CNN ist dabei nicht nur in der Lage, Objekte in Bildern präzise zu klassifizieren, sondern auch diese in einzelnen Instanzen zu segmentieren, was es zu einem mächtigen Werkzeug in Anwendungen wie der medizinischen Bildverarbeitung macht.[4]

3 Materialien

Die für diese Arbeit relevanten Module und Programme wurden in der Programmiersprache Python verfasst. Diese bietet durch ihre einfache Syntax sowie große Anzahl an Bibliotheken für neuronale Netze viele Vorteile für die Untersuchungen dieser Arbeit.

3.1 Verwendete Module

Die für diese Arbeit relevanten Module und Programme wurden in der Programmiersprache Python verfasst. Diese bietet durch ihre einfache Syntax sowie große Anzahl an Bibliotheken für neuronale Netze viele Vorteile für die Untersuchungen dieser Arbeit. Für die anschließende Implementierung des DicomViewers, von dem die Pneumonia-Detection ein Teil ist, findet in C++ statt. Diese Entscheidung gewährleistet eine maximale Effizienz und Performance in der Ausführung, da C++ direkten Zugriff auf Hardware-Ressourcen bietet und sich hervorragend zur Implementierung rechenintensiver Anwendungen eignet.

In dieser Arbeit zur Entwicklung eines Tools zur Pneumonieerkennung auf Röntgenbildern wurden die beiden führenden KI-Frameworks PyTorch und TensorFlow genutzt. Beide Frameworks bieten spezifische Vorteile, die gezielt in verschiedenen Phasen des Projektes eingesetzt werden, um optimale Ergebnisse zu erzielen.

PyTorch bietet eine dynamische Graph-Definition und verbindet diese mit einer intuitiven Handhabung. Daher ist es besonders für die Forschungs- und Experimentierphase geeignet. Somit wird es für die folgenden Untersuchungen der verschiedenen Netzwerke verwendet. Des Weiteren können in dieser Mask R-CNNs besser umgesetzt werden.

Die Stärken von TensorFlow hingegen liegen in seiner guten Skalierbarkeit, sowie effizient in Produktionsumgebungen integriert zu werden. Daher kommt das Modul bei der Vorbereitung der Produktion sowie

für die Optimierung zum Einsatz.

Durch die Kombination von PyTorch und Tensorflow in der Entwicklungsphase und der abschließenden Umsetzung in C++ stellen wir sicher, dass unser Tool sowohl präzise als auch performant ist und den Anforderungen im klinischen Einsatz gerecht wird.

3.2 Datensätze

Die Datensätze, mit denen das Netz trainiert wird, haben einen großen Einfluss auf die Genauigkeit. Hierbei ist besonders relevant, dass der Datensatz groß ist und eine große Masse abdeckt, also auch Sonderfälle enthält. Des Weiteren ist es wichtig, dass die Sonderfälle statistisch richtig dargestellt sind, bzw. alle verschiedenen Typen, z.B. männlich oder weiblich, ausgeglichen vorliegen. Anderenfalls können Phänomene, ähnlich Vorurteilen beim Menschen, auftreten. Da in dieser Arbeit das überwachte Lernen angewandt wird, ist es überdies wichtig, dass der Datensatz vertrauenswürdig ist. Das bedeutet, dass das Labeling von Experten vorgenommen und überprüft wurde, damit das Netz keine Fehler anlernt. Für diese Arbeit werden Röntgenbilder der menschlichen Brust benötigt. Für die Klassifizierung, sowie die Instance Segmentierung von Pneumonie ist überdies das Vorhandensein von Röntgenbildern infizierter Patienten notwendig. Recherchen ergaben zwei passende Datensätze für diese Voraussetzung.

Zum einen den Pediatric Pneumonia Chest X-ray Datensatz., welcher nur die Label ohne Koordinaten der Merkmale enthält ist damit nur eine Klassifizierung nicht jedoch eine Instance Segmentierung möglich. Der Datensatz umfasst 5856 Bilder des Brustkorbs, welche als „Pneumonie“ oder „Gesund“ gekennzeichnet sind. Des Weiteren wird bei der Ursache der Pneumonie zwischen „bakteriell“ und „viral“ unterschieden. Dabei wurden alle Bilder von mindestens zwei Fachärzten klassifiziert[22]. Folgende Recherchen ergaben einen weiteren Datensatz mit insgesamt 30227 Patientenbildern. Dieser besitzt auch Informationen zum Alter sowie Geschlecht der Patienten. Zusätzlich sind auch Koordinaten der Bereiche angegeben, an denen Merkmale einer Pneumonie lokalisiert wurden (Bounding Boxen). Der Datensatz ist in drei Klassen unterteilt NO Lung Opacity/Not Normal; Normal; Lungopacity(Pneumonia), wobei der Anteil an Pneumoniebildern nahezu ein Drittel beträgt.

Des weiteren wurden Datensätze verwendet, welche sowohl aus Röntgenbildern der Lunge, als auch Lungenmasken³. Diese Lerndaten sind die Zusammenführung mehrerer Datensätze. Zu diesen zählen der Montgomery County X-ray Set und JSRT (Japanese Society of Radiological Technology) Database.

Um die Datensätze zu erweitern und das Netz resistenter gegenüber Bildfehlern wie Dehnung, Drehung usw. zu machen, wurde der Datensatz vervielfältigt. Mit Modulen aus der Bibliothek TensorFlow ist es möglich, die Bilder durch Drehen, Dehnen, Verschieben oder Abschwächen des Kontrastes zu verändern, was auch als Data Augmentation bekannt ist. Dadurch können aus einem Bild 2 bis 4 neue Bilder gewonnen werden. Mit Hilfe dieser Methoden war es möglich, einen größeren Datensatz zu erhalten, mit dem die Netze trainiert werden konnten.[27]

4 Vorgehensweise

Im Folgenden wird die Vorgehensweise und die angewendeten Methoden der Arbeit beschrieben, um das Ziel eines möglichst gut angepassten Netzes zu erreichen. Das zu untersuchende Problem dieser Arbeit gehört der Bildklassifizierung sowie der Objectdetection an. Wie in Kapitel 3 (Hintergrund und theoretische Grundlage) beschrieben, besitzt die Convolutional Neural Netzstruktur Vorteile und spezielle Features für solche Probleme. Daher werden diese in der vorliegenden Arbeit verwendet, um eine

³Lungenmasken: binäre Masken, welche jedem Pixel des Röntgenbildes entweder Hintergrund oder Lunge zuordnen

künstliche Intelligenz für die Erkennung von Pneumonie anhand eines Röntgenbildes des Brustkorbes zu erstellen.

4.1 Erstellen eines einfachen CNN

Um das benötigte Netz zu erstellen, wurde im ersten Schritt dieser Arbeit ein eigenes Paket für diese Aufgabe entwickelt. Das Paket „Neural“ besteht aus verschiedenen Modulen und Klassen für die einzelnen Schichten und Funktionen des neuronalen Netzwerkes und dessen Lernprozesses. Das Ziel dieses Paketes ist es, auf möglichst wenige vordefinierte Module zurückzugreifen. Zwingend notwendig ist für das entwickelte Paket nur das Modul „NumPy“, welche fundamentale Rechenprozesse implementiert. Dadurch wurde zum einen ein besseres Verständnis für den Aufbau und die Funktionsweise von CNNs geschaffen, zum anderen ergeben sich dadurch teilweise Möglichkeiten zur Personalisierung. Des weiteren ist die geringe Größe unseres Modules vorteilhaft für kleiner Mikroprozessoren. Dies führt zu einer kostengünstigen Umsetzung der KI und ist daher auch für kleinere Arztpraxen, sowie Gebiete mit schlecht ausgebauter medizinischer Infrastruktur.

Das Paket besteht dabei aus drei Untermodulen. Das erste Modul „Rectifier“ besitzt Klassen für Teilaufgaben, die während des Trainings und der Verwendung des neuronalen Netzes benötigt werden. So sind hier unter anderem die Aktivierungsfunktionen, Optimierer und Initialisierer (für die Parameter des Netzes) definiert.

Des Weiteren besitzt unser Paket noch das Modul „Layer“. Wie der Name aussagt, sind hier die Klassen für die einzelnen Schichten/Layer definiert. Somit können die in Kapitel 2 zur Theorie eingefügten erwähnten Schichten als einzelne Objekte initialisiert und personalisiert werden. Das dritte Modul „Network“ ist in der Lage das ein Netz zu erstellen. In diesem sind die wichtigen Funktionen zum Trainieren und Anwenden des Netzes implementiert.

Mit diesem selbstgeschriebenen Paket war es möglich eine einfaches CNN (ähnlich der LeNet [8] Struktur) zu erstellen.

Im Folgendem wurden die Einflüsse verschiedener Netzparameter auf die Genauigkeit untersucht, um die optimale Netzstruktur für eine präzise Diagnose zu finden. Der Fokus lag dabei auf der Anzahl der Convolutional-Layer und der Epochen, da sie den größten Einfluss auf die Genauigkeit haben. Während viele Schichten komplexere Features extrahieren können, führen zu viele Schichten bei einfachen Problemen zu schlechteren Ergebnissen und erhöhter Rechenlast. Zu wenige Schichten hingegen können zu einer Unteranpassung (Underfitting) führen. Zur Untersuchung wurde ein Programm entwickelt, das das Netz schrittweise um Convolutional- und Pooling-Schichten erweitert und über mehrere Epochen trainiert. Nach jeder Epoche wurde die Genauigkeit anhand von Validierungsdaten überprüft. Die besten Ergebnisse wurden durch die Kombination von Epochenanzahl und Schichtenzahl ermittelt. Zusätzlich kamen Batch- und Dropoutlayer zum Einsatz, um den Lernprozess zu beschleunigen und Overfitting zu verhindern. Des weiteren wurden verschiedene Initialisierer, Optimierer sowie Fehlfunktionen getestet.

4.2 Untersuchungen zur Klassifizierung

Im zweiten Schritt wurden verschiedenen Standardstrukturen für die Bildklassifizierung ausprobiert (siehe Kapitel 2). Die hierfür verwendeten Strukturen sind ResNet, VGG, ResNext, sowie DensNet. Diese lieferten bei ähnlichen Klasifizierungsaufgaben sehr gut Ergebnisse, weshalb diese mithilfe von Transferlearning⁴ an das Problem angepasst werden. Hierfür werden Beide vorgestellten Datensätze für die Bildklassifizierung zusammengelegt. Anschließend werden die Modelle miteinander verglichen. Vergleichskategorien sind hierbei der loss bzw. die Genauigkeit, aber auch die Dauer, welche das Netz für

⁴Transferlearning: Ein bereits an einem sehr großen Datensatz trainiertes Modell wird für eine andere Aufgabe umtrainiert

die Klassifizierung Benötigt. Bei der Implementierung wird für die Untersuchung PyTorch verwendet. Hier sind die Netze, auch bereits mit vortrainierten Parametern, definiert.

Gleichzeitig werden auch Netze nur auf dem Pediatric Pneumonia Chet X-ray Datensatz trainiert. Hier ist die zusätzliche Unterscheidung zwischen bakterieller und viraler Pneumonie möglich.

Um die Entscheidung des Netzes Transparent zu gestalten und zu untersuchen, welche Bereiche des Bildes Maßgeblich für die Entscheidung des Netzes sind, werden weitere Untersuchungen angefügt. Hierbei wird über das Grad-CAM (sowie CAM++ ...) Verfahren diese besonders aktiven Bereiche anhand der letzten Feature-Maps herausgefiltert. Dies kann sowohl über etwaige Merkmale der Pneumonie informieren, als auch auf typische Bereiche einer Pneumonie hinweisen. Des Weiteren ermöglicht dieses Verfahren etwaige Fehlinterpretationen des Netzes zu verstehen. Nicht zuletzt fördern diese Untersuchungen die Transparenz der Netze, was zur Akzeptanz beitragen kann.

4.3 Untersuchung zur Object Detection

Um die Interaktion zwischen Arzt und KI zu verbessern, reicht eine reine Klassifizierung oft nicht aus. Für Ärzte ist insbesondere die Transparenz bei der Diagnoseerstellung durch die KI von großer Bedeutung. Zwar können Grad-CAM-Verfahren zur Sichtbarmachung von Fehlern beitragen und die Transparenz erhöhen, jedoch sind sie oft zu ungenau, um spezifische Merkmale zu identifizieren. Eine effektivere Herangehensweise wäre, dem Arzt nicht nur eine Diagnose zu liefern, sondern auch auffällige Bereiche bzw. Merkmale oder Symptome im Bild zu markieren. Dies könnte die Nachvollziehbarkeit der Diagnose erheblich erleichtern.

Eine solche Verbesserung lässt sich durch Verfahren der Objekterkennung erzielen, die darauf abzielen, Objekte im Bild vom Hintergrund zu unterscheiden und sie zu klassifizieren. Wie in Kapitel 2 erläutert, sind Mask R-CNNs eine weit verbreitete Architektur, die eine hohe Genauigkeit in der Bildanalyse verspricht. Daher wird in diesem Abschnitt der Arbeit ein Mask R-CNN implementiert und mit anderen Architekturen, die aktuell in der Forschung diskutiert werden, verglichen. Zudem wird untersucht, ob eine Kombination von Single-Shot-Detektoren ähnliche Ergebnisse erzielen kann.

4.4 Untersuchung 2-teiliges Netz

Um die Genauigkeit zu steigern wird folgend ein zweigeteiltes Netz untersucht und anhand der bestehenden Netze evaluiert. Dafür wird im ersten Teil ein Netz zur Segmentierung der Lunge erstellt. Ziel dieses Schrittes ist es irrelevante Informationen aus den Bildern zu entfernen und somit sowohl Störfaktoren sowie eine hohe Verarbeitungszeit zu vermeiden. Hierfür werden verschiedenen Architekturen zur Segmentierung untersucht. Maßgeblich sind UNet, sowie deren Variationen, Gated SCNNs und Deeplab Teil der Untersuchung. Hierfür wird der vorgestellte Datensatz mit Röntgenbildern und Lungenmasken verwendet. Nachdem das die beste Architektur ermittelt wurde, werden die vorigen Untersuchung mit den jeweils besten Netzstrukturen wiederholt. So ist eine Evaluierung des vorgestellten 2-Teiligen Netzes möglich.

4.5 Untersuchung kombinierter Netzwerke

Wenn es um das Risiko von Lungenentzündungen geht, müssen verschiedene Faktoren berücksichtigt werden, da sie sich gegenseitig beeinflussen. Ärzte haben bereits den Vorteil diese Hintergrundinformationen über den Patienten in die Diagnose mit einfließen zu lassen. Der Künstlichen Intelligenz hingegen, welche bisher einen auf das Röntgenbild beschränkten Blickwinkel besaß, ist dies nicht möglich.

Das Alter spielt dabei eine große Rolle. Ältere Menschen haben oft ein schwächeres Immunsystem, was bedeutet, dass ihr Körper nicht mehr so gut gegen Infektionen/Entzündungen kämpfen und sie anfälliger gegen Pneumonien sind (siehe linkes Bild) kann. Aufgrund des Datenschutzes ist es nicht möglich, einen

Datensatz zu finden, in welchem Bilder bereits mit allen einer Hintergrundinformation gelabelt sind, welche auch der Arzt besitzt. Für das Testen dieser Erweiterung haben wir uns vorerst auf das Alter als eine Information fokussiert. Mithilfe einer Statistik⁵ war es im Folgenden möglich, verschiedene Alter zu generieren. Dazu wurde anfangs eine Normalverteilung Funktion verwendet, welche die Statistik widerspiegelt. Der Höhepunkt der Statistik liegt hier bei einem Alter zwischen 90 und 110. Später wurde auch noch eine expotentielle Funktion approximiert Je jünger die Altersgruppe wird, desto geringer ist auch der Anteil der Patienten. Aus dieser exponentialen Verteilung werden zufällig so viele Alter entnommen, wie es erkrankte Lungenröntgenbilder gibt. Dabei liegen die meisten bei 110 und die wenigsten bei 0. Anschließend wird jedem erkrankten Röntgenbild unabhängig von der Art der Pneumonie ein Alter zugewiesen. Dasselbe wird anschließend mit den gesunden Röntgenbildern wiederholt, wobei hier der Hochpunkt der Normalverteilung bei 0 und der Tiefpunkt bei 110 zu finden ist. Durch diese Schritte war es möglich einen Datensatz zu konstruieren, in welchem die Röntgenbilder mit einem Alter gelabelt sind. Um die KI der Genauigkeit und Vorteilen eines Arztes anzupassen, wird im Folgenden das Netz um weitere Dateninputs erweitert. Die erste Hürde dieser Idee ist die Übergabe der Daten an das Netz. Da in dieser Arbeit ein Convolutional Neural Network genutzt wird, welches eine mehrdimensionale Eingabe benötigt, können die Daten nicht zusammen mit dem Röntgenbild übergeben werden. Eine Möglichkeit ist es, dem Bild eine Pixelreihe hinzuzufügen, jedoch müssten dann alle Lernbilder über eine Funktion in eine neue Matrix überführt werden. Dies könnte die Interpretation verfälschen und ist für das Netz schwerer zu erfassen. Deshalb bedienen wir uns im Folgenden einem gesplitteten neuronalen Netz. Der erste Teil ist eine Kombination aus Convolutional- und Pooling-Layern für das Röntgenbild. Daneben steht der zweite Teil, welcher eine Aneinanderreihung mehrerer Dense-Layer passend für die eindimensionalen zusätzlichen Patientendaten ist. Beide Teile laufen anschließend in eine Reihe weiterer Dense-Layer zusammen, wo die zuvor gefilterten und bewerteten Daten ausgewertet werden. Dieser Schlussteil fügt beide Inputs zusammen und entscheidet daraus, ob es sich um eine Pneumonie handelt. Da dies jedoch nicht mit unserem Modul möglich war, wird diese Erweiterung abermals mithilfe des Moduls TensorFlow implementiert. Tensorflow stellt Layer für das Zusammenfügen von zwei Netzen bereit⁶. Zusätzlich kann man in der rechten Statistik erkennen dass, ältere Menschen oft an Gewicht zunehmen, was auch Einfluss auf das Immunsystem haben kann. Das liegt daran, dass der Stoffwechsel im Alter langsamer wird und der Körper, aufgrund von Muskelabbau weniger Energie benötigt, was dazu führen kann, dass überschüssige Kalorien als Fett angesetzt werden. Dies kann zu Übergewicht oder sogar Adipositas führen, was wiederum das Risiko für Diabetes erhöhen und das Immunsystem weiter schwächen kann.

⁵Quelle: Adrian Gillissen, Typische Lungenerkrankungen im Alter [23]

⁶Ujjwal Khandelwal, CNN model using Tensorflow - Keras [26]

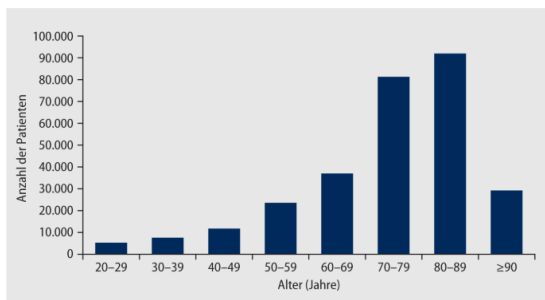


Abbildung 5: BMI-Altersverteilung aus 2017

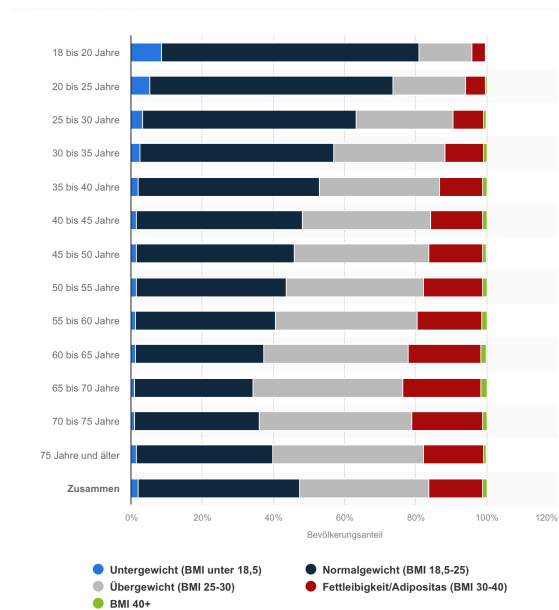


Abbildung 6: Altersverteilung Pneumonie-Erkrankter in Deutschland

Ein weiterer wichtiger Faktor ist das Rauchverhalten. Rauchen ist bekanntermaßen schädlich für die Gesundheit, insbesondere für die Lunge. Es schädigt die Atemwege und schwächt das Immunsystem. Raucher haben auch ein höheres Risiko für Krankheiten wie COPD oder Krebs, die das Immunsystem zusätzlich belasten können. Dabei zeigen Studien der deutschen Krebsgesellschaft dass die Patienten einem höheren Risiko für Krebs aber auch Pneumonien ausgesetzt sind.

Zusammengefasst sind das Alter, der BMI, das Rauchverhalten und andere Krankheiten miteinander verknüpft und beeinflussen sich gegenseitig. Das führt zu dem Schluss, dass je mehr Informationen über die Patienten bekannt sind, desto besser kann der Zustand und die Diagnose des Patienten eingeschätzt und Vorkehrungen getroffen werden.^{7 8 9} Die Untersuchungen zu dem Alter wurden mit dem zweiten Datensatz in welchem dieses vorhanden ist wiederholt und mit der vorigen verglichen. Folglich wurden für den BMI und das Rauchverhalten ähnlich wie für das Alter Verteilungen aufgesetzt und diese dem Datensatz zugewiesen. Des Weiteren banden wir auch weitere Diagnoseverfahren ein. CT-Scans bieten eine höhere Auflösung als Röntgenbilder und so sind auch kleiner Merkmale auf diesen oft deutlicher. Deshalb banden wir zusätzlich auch ein weiteres CNN in das gesplittete Netzwerk ein, welches die CT-Scans analysiert. Um das Netz zu trainieren, passten wir die Gewichte des CNNs für die CT-Scans erst extern an einen Datensatz mit rund 824 Bildern (welchen wir auf 2000 Bilder erweiterten) an. Dieses Netz wurde anschließend in das gesplittete Netz als zusätzlichen Teil eingebunden. Das so entstandene Netz wurde noch einmal zusammen trainiert, wobei den Röntgenbildern zufällige CT-Scans zugewiesen wurden. Dabei wurden nun die Parameter des CT-Teils eingefroren und ausschließlich die anderen Parameter angepasst, um ein Overfitting beim CT-Teil zu vermeiden.

⁷Klaus Dalhoff, Ambulant erworbene Pneumonie bei Erwachsenen [21]

⁸Fr. Dr. Daniela M. Christmann, Rauchen - Zahlen und Fakten [20]

⁹Rainer Radtke, Anteile an der deutschen Erwachsenenbevölkerung nach Körpermaßen (BMI)* und Altersgruppen im Jahr 2017 [29]

4.6 Integration der KI in den medizinischen Workflow

Um das Tool für Ärzte verfügbar zu machen wurde zusätzlich die Integration dieser KI in den Medizinischen Workflow untersucht. Dazu wurde mit der Hilfe von Fachärzten des Universitätsklinikum ein Schema entworfen. Des Weiteren wurde ein DicomViewer¹⁰ um dieses Schema zu Realisieren. Dafür wurde in C++ eine Graphische Oberfläche entwickelt mit Standardfunktionen

Das finale Ziel unseres Projekts ist die Entwicklung einer effizienten und schnellen Anwendung für den PC, die Pneumonie auf Röntgenbildern erkennt. Während in der Entwicklungs- und Trainingsphase PyTorch und TensorFlow verwendet werden, erfolgt die finale Implementierung in C++. Diese Entscheidung gewährleistet eine maximale Effizienz und Performance, da C++ direkten Zugriff auf Hardware-Ressourcen bietet und sich hervorragend zur Implementierung rechenintensiver Anwendungen eignet.

Durch die Kombination von PyTorch und TensorFlow in der Entwicklungsphase und der abschließenden Umsetzung in C++ stellen wir sicher, dass unser Tool sowohl präzise als auch performant ist und den Anforderungen im klinischen Einsatz gerecht wird. Um das Tool für Ärzte verfügbar zu machen, wurde zusätzlich die Integration dieser KI in den medizinischen Workflow untersucht. In Zusammenarbeit mit Fachärzten des Universitätsklinikums Jena wurde ein Schema entwickelt, das die Einbindung in die klinische Praxis erleichtert. Zur Realisierung dieses Schemas wurde ein DicomViewer¹¹ entwickelt, der eine grafische Oberfläche mit Standardfunktionen bisheriger medizinischer Viewer bietet. Dieser Viewer dient als Schnittstelle zwischen der KI und dem Arzt, sodass die Anwendung direkt im klinischen Alltag getestet und eingesetzt werden kann.

Darüber hinaus wurde das Programm so modifiziert, dass es in Zukunft auch vor Ort in Arztpraxen oder anderen medizinischen Einrichtungen lokal trainiert werden kann. Auf diese Weise verlassen die Patientendaten niemals die Praxis, und lediglich die Parameter des trainierten Netzes werden an eine zentrale Stelle zurückgesendet. Diese Vorgehensweise ermöglicht es, den Datensatz kontinuierlich zu erweitern, ohne Datenschutzbestimmungen zu verletzen.

Abschließend wird das entwickelte Tool in der realen Umgebung des Universitätsklinikums Jena getestet, um sicherzustellen, dass es den hohen Anforderungen im medizinischen Bereich gerecht wird und eine zuverlässige Unterstützung bei der Diagnose von Pneumonie bietet.

Programme und Datensätze

Die Programme und verwendeten Datensätze wurden so weit wie möglich auf Github hochgeladen. Unter dem folgenden Link: <https://github.com/Hey1Marvin/PneumoniaJuFo> kann auf diese Einsicht genommen werden. Jedoch war es uns nur möglich, die Hauptprogramme und einen gekürzten Datensatz hochzuladen, da uns keine Server für die benötigte Datenmenge zur Verfügung stehen. Wir arbeiten noch daran alle ordentlich dokumentiert auf der beschriebenen Seite hochzuladen. Unser ursprünglicher Datensatz stammt dabei von der Website huggingface.com, wir konnten allerdings noch einen weiteren Datensatz von der NIH (National Institutes of Health) finden. Dabei gab es schon einen Datensatz nur für Pneumonia mit 30000 Bildern. Zusätzlich enthält dieser Datensatz Bilder mit Boundingboxen, welche perfekt zum Trainieren des Mask R-CNN und damit der Fähigkeit des Netzes, erkannte Stellen in den Bildern zu umkreisen, war.

¹⁰DicomViewer: Image Viewer für Medizinische Bilder (häufig im dicom Format)

¹¹DicomViewer: Image Viewer für Medizinische Bilder (häufig im dicom Format)

5 Literaturverzeichnis

Gedruckte Literatur

- [1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv preprint arXiv:1606.00915*, 2016.
- [2] Foivos I Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. Resunet-a: a deep learning framework for semantic segmentation of remotely sensed data. *arXiv preprint arXiv:1904.00592*, 2019.
- [3] Dirk Hecker Gerhard Paaß. *Künstliche Inteligenz*. Springer Vieweg, Braunschweig/Wiesbaden, 2020.
- [4] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [6] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.
- [7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems (NIPS)*, pages 1097–1105, 2012.
- [8] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [9] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.
- [10] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 21–37, 2016.
- [11] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016.
- [12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241, 2015.
- [13] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*, 2015.
- [14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

- [15] Towaki Takikawa, David Acuna, Varun Jampani, and Sanja Fidler. Gated-scnn: Gated shape cnns for semantic segmentation. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 5229–5238, 2019.
- [16] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1492–1500, 2017.
- [17] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11, 2018.

Internetliteratur

- [18] Eli Bendersky. Backpropagation through a fully-connected layer.
<https://eli.thegreenplace.net/2018/backpropagation-through-a-fully-connected-layer/>,
Zugriffsdatum: 13.02.2023.
- [19] Jason Brownlee. How do convolutional layers work in deep learning neural networks?
<https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/>,
Zugriffsdatum: 28.12.2022.
- [20] Fr. Dr. Daniela M. Christmann. Rauchen - Zahlen und Fakten.
<https://www.krebsgesellschaft.de/onko-internetportal/basis-informationen-krebs/bewusst-leben/rauchen-zahlen-und-fakten.html>, Zugriffsdatum: 19.04.2024.
- [21] Klaus Dalhoff. Ambulant erworbene Pneumonie bei Erwachsenen.
<https://link.springer.com/article/10.1007/s11298-017-5989-y>, Zugriffsdatum: 19.04.2024.
- [22] EMLÅM1998. chest-xray-Pneumonia.
<https://www.kaggle.com/datasets/emlm1998/chestxraypneumonia>, Zugriffsdatum: 14.12.2023.
- [23] Adrian Gillissen. Typische Lungenerkrankungen im Alter.
<https://doi.org/10.1007/s42090-020-0622-8>, Zugriffsdatum: 13.12.2023.
- [24] Jeremy Jordan. Common architectures in convolutional neural networks.
<https://www.jeremyjordan.me/convnet-architectures/>, Zugriffsdatum: 25.08.2024.
- [25] MathWorks. Was ist ein convolutional neural network?
<https://de.mathworks.com/discovery/convolutional-neural-network.html>, Zugriffsdatum: 25.08.2024.
- [26] CNN model using Tensorflow Keras. Ujjwal khandelwal.
https://pythonandml.github.io/dlbook/content/convolutional_neural_networks/cnn_architecture.html, Zugriffsdatum: 22.11.2023.
- [27] Paul Mooney. Chest X-Ray Images (Pneumonia).
<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>, Zugriffsdatum: 09.08.2022.
- [28] Julia Odenthal. Convolutional neural networks.
<https://www.mi.uni-koeln.de/wp-znikolic/wp-content/uploads/2019/06/11-Odenthal.pdf>, Zugriffsdatum: 27.12.2022.

- [29] Rainer Randtke. Anteile an der deutschen Erwachsenenbevölkerung nach Körpermaßen (BMI)* und Altersgruppen im Jahr 2017.
<https://de.statista.com/statistik/daten/studie/256578/umfrage/bevoelkerungsanteile-in-deutschland-nach-koerpermassen-bmi-und-altersgruppen/>,
Zugriffsdatum: 19.04.2024.
- [30] Savyakhosla. Cnn | introduction to pooling layer.
<https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/>, Zugriffsdatum:
03.01.2022.
- [31] Max-Ludwig Stadler. Convolutional neural network (cnn).
<https://tinyurl.com/ytc78n5d>, Zugriffsdatum: 26.12.2022.
- [32] Sakshi Wani, Insha; Arora. Osteoporosis diagnosis in knee x-rays by transfer learning based on convolution neural network.
https://www.researchgate.net/figure/Basic-Network-architecture-of-AlexNet-ResNet-18-VggNet-19-and-fig3_363798567, Zugriffsdatum: 25.08.2024.
- [33] Thomas Wood. Convolutional neural network.
<https://deepai.org/machine-learning-glossary-and-terms/convolutional-neural-network>,
Zugriffsdatum: 27.12.2022.

Eidesstattliche Erklärung

Wir erklären eidesstattlich, dass wir die Arbeit selbständig angefertigt, keine anderen als die angegebenen Hilfsmittel benutzt und alle aus ungedruckten Quellen, gedruckter Literatur oder aus dem Internet im Wortlaut oder im wesentlichen Inhalt übernommenen Formulierungen und Konzepte gemäß den Richtlinien wissenschaftlicher Arbeiten zitiert, durch Fußnoten gekennzeichnet oder mit genauer Quellenangabe kenntlich gemacht haben.